

方向変換モデルを用いた 複数カメラ間での人物照合

川合 諒^{1,a)} 楨原 靖^{1,b)} 村松 大吾^{1,c)} 岩間 晴之^{1,d)} 八木 康史^{1,e)}

概要: 本研究では、歩行者のカメラ渡りの追跡を目的とする、方向変換モデルを用いた複数カメラ間での人物照合手法を提案する。特徴には、時空間輝度勾配に基づく歩容特徴である STHOG 特徴を用いる。特徴は、背景部分のエッジを弱めた上で、異なる観測方向の対応領域の高さが合うように、カメラ姿勢と歩行方向の推定を基に変形させたウィンドウから抽出する。方向変換モデルの学習段階では、様々な観測方向から撮影した認識対象外の歩行画像列を収集し、それぞれの歩行姿勢（位相）を同期させた規定枚数の部分画像列から STHOG 特徴列を抽出する。そして、各行が観測方向、各列が各被験者の各歩行姿勢と対応するように特徴を並べた行列を作成する。その行列を特異値分解することにより、方向を変換する行列を得る。認証段階では、観測方向の異なる二つの歩行画像列が与えられると、方向変換モデルを用いて、一方の方向の STHOG 特徴列をもう一方の方向に変換し、同一方向の下で位相同期を行いながら照合する。実験では、大学で撮影した歩容映像を用いて人物照合を行い、提案手法が有用であることを確認した。

1. はじめに

近年、人々の防犯意識の高まりに伴い、防犯カメラが急速に普及している。商店や銀行といった施設内だけでなく、商店街や市街地の道路などの公共の場所、更には個人宅に設置される事例も増加している。防犯カメラには、犯罪の瞬間や証拠の記録、監視されている事による犯罪の抑止といった機能が考えられるが、何らかの事件が起こった際、現場周辺の防犯カメラを確認し、不審人物がいないか検索するという用途もある。

その際には、不審人物は一定の場所に留まるようなことはなく、複数のカメラを横断して移動すると考えられる。その人物を追跡することによって、不審人物の住所など、詳細な情報が分かり得る。最近、この手法によって容疑者が割り出される事件や、特別手配犯を追跡して確保へつながる事例などが増えている。しかし、不審人物の追跡を人手によって行った場合、カメラの組み合わせが膨大になることから、極めて多くの時間を要することになる。よって、捜査経済の観点から、自動的にカメラ渡りの人物追跡を行うことが重要となる。

異なるカメラ間での人物照合の手法としては、色やテクスチャを用いるものが提案されてきた。どちらも、カメラから遠く離れた人物に対しても認証可能な手段であるが、他に似た色の服を着ている人がいる場合に誤認識する可能性があるなどの問題がある。

歩容認証は、人物の歩き方から個人を識別する認証方法である。形状特徴に加え、姿勢の違いをも個人の特徴として利用して認証するため、認証精度の向上が期待できる。しかし、歩容認証には、人物の観測角の違いにより、特徴が大きく異なる可能性があるという問題がある。

そこで、本研究では、歩容特徴を方向変換モデル (VTM) を用いて変換し、その特徴を用いて複数カメラ間での人物照合を行うことを提案する。歩容特徴を取得する際は、カメラ姿勢と歩行方向推定から、3次元的に人物領域を指定する事を考える。特徴量としては、時空間画像におけるエッジの方向と強度から作成したヒストグラムである STHOG 特徴を用いる。

本論文の構成は以下の通りである。2章で人物照合についての関連研究を紹介し、3~7章で提案手法を説明する。この中で、3章で研究の前提条件、4章で提案手法の概要、5で特徴抽出、6章で方向変換、7章で実際の照合について説明する。8章において実際の歩容画像を用いた実験の説明と結果の考察を行い、9章でまとめと今後の課題について述べる。

¹ 大阪大学
Osaka University

a) kawai@am.sanken.osaka-u.ac.jp

b) makihara@am.sanken.osaka-u.ac.jp

c) muramatsu@am.sanken.osaka-u.ac.jp

d) iwama@am.sanken.osaka-u.ac.jp

e) yagi@am.sanken.osaka-u.ac.jp

2. 関連研究

ここでは、複数カメラ間の人物照合手法のうち、特にカメラからの見え方に基づく従来手法について述べる。

2.1 色に基づく人物照合

現在提案されている複数カメラ間での人物照合において、非常に多くの手法で利用されているのが色特徴である。色特徴としては、ヒストグラムが多く用いられる [1][2][3][4][5]。特に、Cheng ら [4] や Cai ら [5] は、一定幅でなく特徴的な色や生起頻度が高い色をピンとするヒストグラムを使用していることが特筆される。ヒストグラムは一般に空間的特徴を無視するが、Bird ら [6] は人物領域の分割、Lin ら [7] は色空間と高さ方向のジョイントヒストグラムの作成により、それぞれ空間的な情報を考慮した色特徴を提案している。また、各カメラでホワイトバランスや明るさ等の設定が異なると、本来同一な色の特徴も異なる可能性があるが、同じく [7] では、RGB の各輝度の相対順位を特徴量とすることで違いを吸収することを提案しているほか、そのような状況でのカラーキャリブレーション手法も Porikli によって提案されている [8]。

色特徴は、ある程度人物の観測方向が違っていても、特徴が大きく変化しないという特長がある。しかし、色特徴の問題点として、暗い色のコートを多くの人が着用する冬のように、複数の人物が似た服を着ている場合、特徴の差が小さくなり、認証率が低下するという点が挙げられる。

2.2 テクスチャに基づく人物照合

Hamdoun ら [9] は、特徴点の抽出による動画ベースの人物照合を提案している。また、Bak ら [10] や Viola と Jones [11] は Haar-like 特徴量を用い、Adaboost によって識別器を構築することを提案している。Wang ら [12] は Log をとった RGB 色空間における HOG (Histograms of Oriented Gradient) と Lab 色空間において HOG に対して shape context を計算したものの共起性を特徴量としている。また、Alahi ら [13] や Bak ら [14] は共分散を用いている。共分散は、x 方向と y 方向に関する共起性を参照することで輝度や勾配の空間的な配置を表現する。

色とテクスチャそれぞれ単独では精度が低い場合が多いため、テクスチャや色を組み合わせた手法も多数提案されている。例えば、上村ら [15] は、RGB ヒストグラムとオートコリログラムを組み合わせたことを提案している。また、Gheissari ら [16] は、特徴点ベースの Hessian affine 不変特徴量から人物の手足等の領域毎の対応付けを行っている。Berdugo ら [17] は、Lin らの [7] を拡張し、服装の模様に関する情報を取り出すために ranked color ratio, ranked oriented gradients, ranked saliency map と呼ばれ

る特徴量を用いることを提案している。中でも、Farenzena ら [18] の Symmetric Drove Accumulation of Local Features (SDALF) と呼ばれる手法は、色特徴として重み付き HSV ヒストグラムと Maximally Stable Color Region、模様特徴として Recurrent High-Structured Patches を用い、現状最高レベルの精度の照合精度が得られている。ほか、Bazzani ら [19] は、HSV ヒストグラムと Epitome を利用している。Epitome は、得られた画像を複数のパッチ要素を平均とするガウス混合モデルで表現するものである。

2.3 方向変化を考慮した人物照合

複数カメラ間での人物照合を行うにあたっては、歩行者の観測角が二つのカメラの間で異なることを考慮しないと、特にテクスチャを用いる場合、認証率が低下する。このことを考慮した識別手法がいくつか提案されている。

観測方向が異なる場合、本人同士でも非類似度（距離）が遠くなるため、本人と他人を分ける絶対的な基準を設けることは困難である。そのため、本人間の距離が他人間の距離より相対的に小さくなるようにする必要がある。このような距離関係を学習するため、人物照合問題を、Prosser ら [20] は Ranking 問題に、井尻ら [21] や Zheng ら [22] は、距離指標学習の問題に帰着させることを提案している。

そのほか、Gandhi ら [23] は、カメラと人との幾何関係を基に Panoramic Appearance Map (PAM) と呼ばれる方向によらない画像を作成し、それによる照合を提案している。

3. 前提条件

本研究においては、次に挙げる処理は事前に済んでいるものと仮定する。

- 歩行者検出
STHOG 特徴と AdaBoost による歩行者検出器 [24] を用いる。
- 同一カメラ内の人物追跡
Wang と Yagi による手法 [25], [26], [27] など、従来の追跡手法を利用する。

従って、以下では、与えられた歩行者ウィンドウ系列を異なるカメラ間で照合する問題を対象とする。

4. 提案手法の概要

提案手法では、以下の流れで人物照合を行う。それぞれの内容の詳しい説明は次節以降で行う。

まず、全ての歩行者について、特徴を抽出する。特徴抽出の手順は以下の通りである。

- (1) ウィンドウの設定
- (2) 背景エッジの低減
- (3) STHOG 特徴の計算

次に、得られた特徴から、方向変換を行う。方向変換のフレームワークを図 1 に示す。なお、以下本論文では二

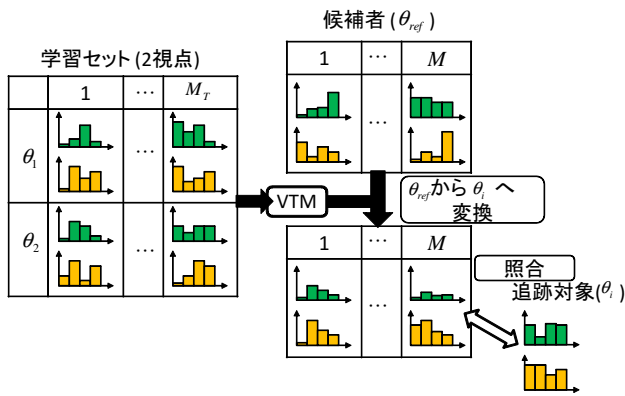


図 1 方向変換モデルのフレームワーク
Fig. 1 Framework of view transformation.

つの視点間の変換について述べる。

方向変換は方向変換モデル [28] によって行われ、大きく学習段階と認証段階に分けられる。

学習段階では、まず、両方の視点からの歩容特徴列を取得する。そして、同一人物の間で位相が一致する部分を取り出し、その部分特徴列と各視点の観測方向の情報をモデル学習に用いる。

認証段階では、歩容特徴と観測方向が画像列から得られる。追跡対象者の観測方向を θ_i 、候補者の観測方向を θ_{ref} とすると、候補者の歩容特徴列それぞれに方向変換モデルが適用され、 θ_i の観測方向に変換された歩容特徴が得られる。

本研究では、方向変換モデルを体の部分ごとに作成することで、少ないデータで効率的に方向の変化を学習できるようにする。

5. 特徴抽出

5.1 ウィンドウの設定

STHOG 特徴では、後述の通り、歩行者検出ウィンドウを細分化したセル単位で特徴を抽出し、照合を行う。そのため、本来であれば、同じセルには同じ体の部位が入っているべきである。しかし、カメラ間で観測方向の変化がある場合、同じセルでも含まれる体の部分が異なってくる。例えば、図 2 は、同じ歩行者のほぼ同じ姿勢の瞬間を異なる視点から撮影したもので、同じ形のウィンドウと足元のセルを当てはめているが、正面に近いカメラでは片足が足元のセルに含まれていないのに対し、もう片方の真横から映しているカメラでは両足とも含まれていることが分かる。

これに対応するため、カメラの姿勢と歩行方向に基づき、3次元的にウィンドウを設定する。キャリブレーションは既存手法を用いる。例えば、構造物から3方向の消失点を求める方法を用いると、図 3 のように3次元座標が定義される。ただし、図上の軸は方向を示すためあくまで便宜的に描いたもので、交点が原点ではないことに注意されたい。

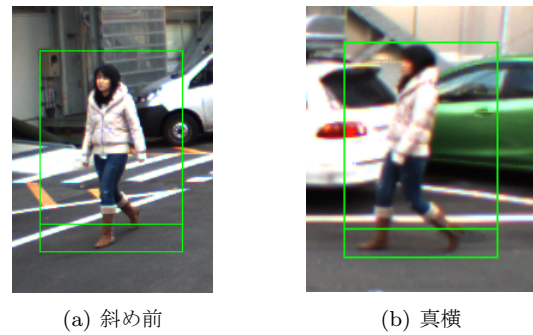


図 2 方向変化によるセルの対応
Fig. 2 Cell correspondence change due to view change.

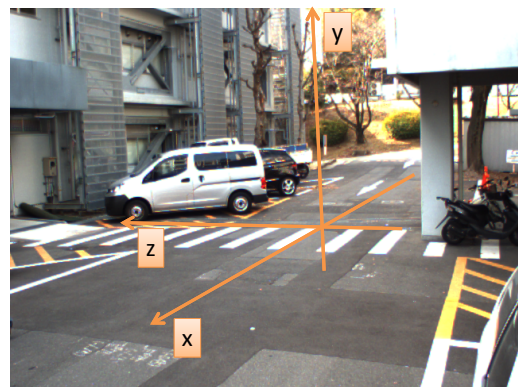


図 3 世界座標における 3 軸の方向
Fig. 3 Orientation of x, y, z axis in world coordinate.



図 4 カメラ姿勢を考慮したウィンドウの設定例
Fig. 4 Window considering camera pose

次に、歩行者の検出ウィンドウにおける足元点の軌跡から3次元的な歩行方向を推定し、その歩行方向が新たな x 軸となるよう、座標を y 軸を中心に回転させる。図 2 の場合、歩行者は x 軸に平行に歩いており、調整の必要はない。そして、歩行方向に応じて、横方向に近い方向で歩いている場合は xy 平面、奥行き方向に近い方向で歩いている場合は yz 平面に平行になるようウィンドウを設定する。すると、図 2 に対応するウィンドウは図 4 のようになり、一番下のセルに両足を含むようになったことがわかる。

5.2 背景エッジの低減

ウィンドウの中には背景が含まれるため、その影響を抑える必要がある。しかし、実世界の映像においては、従来の背景差分による方法では、照明変動等の影響によりシル

エッジ品質が低下することが多い。また、提案手法では、歩容特徴としてエッジの情報を用いており、人物の周囲のエッジは最も重要な情報になるが、シルエット品質が低い背景差分で最も不安定な部分も人物の周囲であり、重要なエッジを安定して得られない可能性がある。

したがって、当研究では、Background Attenuation という手法 [29] を用いて背景のエッジを直接低減させることで、背景の影響を抑えることとする。

この手法の基本的な考え方は、あるピクセルにエッジが存在した場合、同じ場所の背景画像でもエッジが存在し、かつエッジのパターンが背景と入力画像で類似している際に、その部分のエッジを弱める、というものである。これを式で表すと、

$$G_s = \frac{\sqrt{G_x^2 + G_y^2}}{\left(1 + KG_x^{B^2} e^{-\frac{z_x^2}{\sigma^2}}\right) \left(1 + KG_y^{B^2} e^{-\frac{z_y^2}{\sigma^2}}\right)} \quad (1)$$

となる。ここで、 K と σ は Background Attenuation の程度をコントロールするハイパーパラメータであり、 G_x^B と G_y^B は背景の x または y 軸方向の勾配の強さを意味する。 z_x と z_y は背景画像と前景画像との間での x または y 方向のエッジパターンの類似度を示し、

$$z_x = \max\{\|I_{x,y}^B - I_{x,y}\|, \|I_{x+1,y}^B - I_{x+1,y}\|\} \quad (2)$$

$$z_y = \max\{\|I_{x,y}^B - I_{x,y}\|, \|I_{x,y+1}^B - I_{x,y+1}\|\} \quad (3)$$

で定義される。 $I_{x,y}^B$ と $I_{x,y}$ はピクセル (x,y) の背景または前景の強度を表す。

5.3 STHOG 特徴

STHOG (Spatio-Temporal Histograms of Oriented Gradient) 特徴 [24] は、エッジの勾配の強度 $I(x,y) = \sqrt{I_x^2 + I_y^2}$ と方向 $\phi(x,y) = \tan^{-1}(I_y/I_x)$ から作成されたヒストグラムを特徴量とする HOG 特徴 [30] を時空間画像に拡張したものである。

対象の人物が映っている画像のウィンドウをいくつかのセルに分割し、時空間画像から得られるエッジの空間勾配方向 ϕ と時間勾配方向 θ について、セル毎にヒストグラムを作成する。その各ヒストグラムの値をベクトルとして並べたものが、STHOG 特徴における特徴ベクトルである。空間勾配方向は対象とする人物の形状特徴、時間勾配方向は対象とする人物の動き特徴を表している。

特徴量は以下のように計算する。時空間輝度勾配

$$\nabla I = [I_x, I_y, I_t]^T = \left[\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}, \frac{\partial I}{\partial t} \right]^T \quad (4)$$

を用いて、勾配強度 I と、前述の ϕ , θ を式 (5)–(7) のように定義する。

$$I(x,y,t) = \sqrt{I_x^2 + I_y^2 + I_t^2} \quad (5)$$



図 5 歩行者の傾きの補正

Fig. 5 Adjusting leaned pedestrian.

$$\theta(x,y,t) = \tan^{-1} \left(\frac{I_t}{\sqrt{I_x^2 + I_y^2}} \right) \quad (6)$$

$$\phi(x,y,t) = \tan^{-1} \left(\frac{I_y}{I_x} \right) \quad (7)$$

ここで、歩行者の傾きの修正を考える。カメラの配置によっては、歪みの影響で、歩行者が（まっすぐ立っていても）傾いたように画像上に映ることがある（例：図 5）。この場合は、キャリブレーションを基にウィンドウを設定すると、平行四辺形に近い形になる。そこで、ウィンドウの上辺と下辺のそれぞれ中点を結んだ直線の傾きを調べ、その傾きの分だけ空間勾配 ϕ を調整する。図 5 の場合、空間勾配 ϕ の値は式 7 で求めた ϕ から α だけ引くことになる。

その後、 ϕ と θ をそれぞれの階級とし、ピクセル毎のエッジ強度 I を該当する階級に投票することで、一つのセル毎に二つのヒストグラムを作成する。

なお、本研究では、特徴一つを連続する 3 フレームから取得し、ウィンドウを鉛直方向に 7 分割したものを一つのセルとした。また、 $0^\circ \leq \theta < 180^\circ$, $0^\circ \leq \phi < 90^\circ$ を角度の変域とし、9 段階のビンに分けてヒストグラム化した。したがって、一つのウィンドウの特徴量は $9 \times 2 \times 7$ より 126 次元空間で表される。

6. 方向変換

6.1 位相合わせ

方向変換モデルを作成する際は、方向の変化以外の要素をできる限りなくすため、位相を一致させることが必要である。この位相合わせは、後述の照合と似た手順で、以下のように行う。前述のとおり視点は二通りあるため、片方をカメラ 1、もう片方をカメラ 2 とする。

まず、カメラ 1 については、全体のシーケンスから、学習に必要な分の部分シーケンスを、全体の中央寄り取得する。これを基準として、カメラ 2 の位相を合わせる。

カメラ 2 の部分シーケンスも、中央付近で取得できるよう、カメラ 2 の全シーケンスの中で時間的に最も中央に位置する特徴（これを c と表すこととする）を含むようにする。中央付近で取得することで、同一カメラ内での方向変

化をできるだけ小さくする意図がある。

特徴 c がサブシーケンスの最後にある状態から始め、1 フレームずつ右にスライドする。それぞれ、カメラ 1 のシーケンスとこのサブシーケンスの類似度を照合と同様に計算し、 c がサブシーケンスの最初に来るまで繰り返す。最も非類似度が低かったサブシーケンスを、位相が一致したシーケンスとみなす。

6.2 方向変換モデルの学習

前述のとおり、セルごとに別々の方向変換モデルを学習、利用する。

学習段階においては、 K 通りの観測方向から撮影した M 人の特徴を学習データとして利用する。それぞれ、1 人あたり N 個の特徴を学習する。ここで、 k 番目の角度から撮影した m 人目の n 番目の特徴ベクトルを $\mathbf{a}_{\theta_k}^{m,n}$ と定義する。

次に、式 8 に示すように、特徴ベクトルを並べて行列を作成し、特異値分解する。

$$\begin{bmatrix} \mathbf{a}_{\theta_1}^{1,1} & \cdots & \mathbf{a}_{\theta_1}^{1,N} & \cdots & \mathbf{a}_{\theta_1}^{M,1} & \cdots & \mathbf{a}_{\theta_1}^{M,N} \\ \vdots & & & \ddots & & & \vdots \\ \mathbf{a}_{\theta_K}^{1,1} & \cdots & \mathbf{a}_{\theta_K}^{1,N} & \cdots & \mathbf{a}_{\theta_K}^{M,1} & \cdots & \mathbf{a}_{\theta_K}^{M,N} \end{bmatrix} = USV^T = \begin{bmatrix} P_{\theta_1} \\ \vdots \\ P_{\theta_K} \end{bmatrix} \begin{bmatrix} v^{1,1} & \cdots & v^{M,N} \end{bmatrix} \quad (8)$$

6.3 方向変換モデルの適用

認証段階においては、入力を $\mathbf{a}_{\theta_{ref}}^{m,n}$ とすると、

$$\mathbf{a}_{\theta_k}^{m,n} = P_{\theta_k} v^{m,n} \quad (9)$$

が成り立つことから、方向 θ_{ref} から方向 θ_i への変換は、以下のように得られる。

$$\hat{\mathbf{a}}_{\theta_i}^{m,n} = P_{\theta_i} (P_{\theta_{ref}}^T P_{\theta_{ref}})^{-1} P_{\theta_{ref}}^T \mathbf{a}_{\theta_{ref}}^{m,n} \quad (10)$$

7. 照合

c_1 番目のカメラに映った i_1 番目の追跡対象者（以下プローブという）と、 c_2 番目のカメラに映った i_2 番目の候補者（以下ギャラリーという）を例に、照合を行う際の手順について詳しく記す。

STHOG 特徴は、前述した通り、フレーム毎に時空間勾配に基づいて計算されるが、単独フレームでは一つの歩容姿勢の特徴に限定され、十分な識別能力が得られない。よって、複数フレームからなる STHOG 特徴の系列を単位として照合する。

c 番目のカメラに映った i 番目の人物の、 j 番目のフレームに対する特徴を $\mathbf{a}_{i,j}^c$ 、その人物の特徴の総数を J_i^c とし、全特徴、すなわちからまでの特徴の系列を

表 1 大学構内のカメラの仕様

Table 1 Specification of cameras in the university.

機種	IEEE1394 カメラ Flea2 (Point Grey Research 社製)
フレームレート	30fps
解像度	640 × 480 ピクセル

$A_i^c = \{\mathbf{a}_{i,0}^c, \dots, \mathbf{a}_{i,J_i^c-1}^c\}$ とする。また、照合単位のフレーム数を p とし、それを平均的な歩行周期と考えられる 1 秒になるように設定する。 $\mathbf{a}_{i,j}^c$ から $\mathbf{a}_{i,j+p-1}^c$ までの p 個の特徴の系列を $A_{i,j}^c$ とし、各フレームで対応する特徴間のノルムの総和を以下で定義する。なお、ノルムとしては過去の実験 [31] で最良の結果が得られた L0.5 ノルムを用いる。

$$D(A_{i_1,j_1}^{c_1}, A_{i_2,j_2}^{c_2}) = \sum_{t=0}^{p-1} \|\mathbf{a}_{i_1,j_1+t}^{c_1} - \mathbf{a}_{i_2,j_2+t}^{c_2}\| \quad (11)$$

次に、プローブ系列を p フレーム毎の $k = \lfloor J_{i_1}^{c_1}/p \rfloor$ 個の部分系列、 $A_{i_1,0}^{c_1}, A_{i_1,p}^{c_1}, \dots, A_{i_1,(k-1)p}^{c_1}$ に分割する。プローブの各部分系列 $A_{i_1,lp}^{c_1} (0 \leq l < k)$ について、ギャラリー系列と位相シフトしながら比較した場合の最小値、

$$m(A_{i_1,lp}^{c_1}, A_{i_2}^{c_2}) = \min_{0 \leq s \leq J_{i_2}^{c_2} - p} D(A_{i_1,lp}^{c_1}, A_{i_2,s}^{c_2}) \quad (12)$$

を計算する。この計算により、歩容の位相を合わせた、部分系列 $A_{i_1,lp}^{c_1}$ とギャラリーとの距離が得られる。更に、その値に関するプローブの部分系列間での最小値、

$$m(A_{i_1}^{c_1}, A_{i_2}^{c_2}) = \min_{0 \leq l < k} m(A_{i_1,lp}^{c_1}, A_{i_2}^{c_2}) \quad (13)$$

を求め、この $m(A_{i_1}^{c_1}, A_{i_2}^{c_2})$ をプローブとギャラリーとの距離と定義し、その値を非類似度とすることで人物を照合する。即ち、あるプローブを各ギャラリーと比較し、最も距離が小さかったギャラリーと、そのプローブが同一人物であるとみなす。

8. 実験

8.1 データセット

大阪大学産業科学研究所内の道路（以下、大学構内と呼ぶ）において、2m の高さに図 6 のように設置した 3 台のカメラを用いて、27 人の被験者の歩容映像を撮影した。このうち 14 人を方向変換モデルの学習セット、残る 13 人をテストセットとして利用した。

3 台のカメラの仕様は表 1 に示す通りであり、被験者は図 6 における赤色の線上を歩行する。カメラ A,B とカメラ C の間に十分な距離をとっているのは、大きく異なる背景条件下でのデータを収集するためである。また、明るさや色合い等はカメラ間で統一せず、自動ゲイン制御によって撮影した。撮影画像の例を図 7 に示す。

3 台のカメラで歩容を撮影するが、同一カメラ内であっても、画像の右寄りと左寄りとは観測角がある程度異

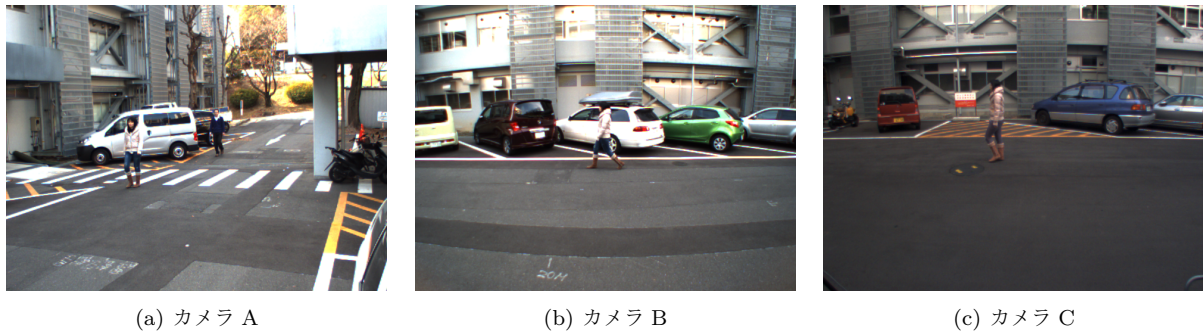


図 7 大学構内の映像の例

Fig. 7 Example of images captured by cameras in the university.

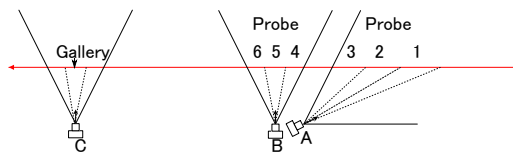


図 6 大学構内のカメラの配置

Fig. 6 Arrangement of cameras in the university.

プローブ						ギャラリー
1	2	3	4	5	6	

図 8 データの例

Fig. 8 Example of data.

なってくる。そのため、一つのカメラの画像シーケンスを三つに分割し(図 6 上の点線)、一人あたり合計 9 シーケンスを得る。カメラ A, B の 6 シーケンスをプローブ、カメラ C の真ん中のシーケンスをギャラリーとして比較する。図 6 中の数字をシーケンス番号として以降の説明に用いる。ギャラリーとプローブのシーケンス 5 の観測角がほぼ同じで、そこから離れるにつれて観測角の差が大きくなる。図 8 にそれぞれのシーケンスの例を示す。

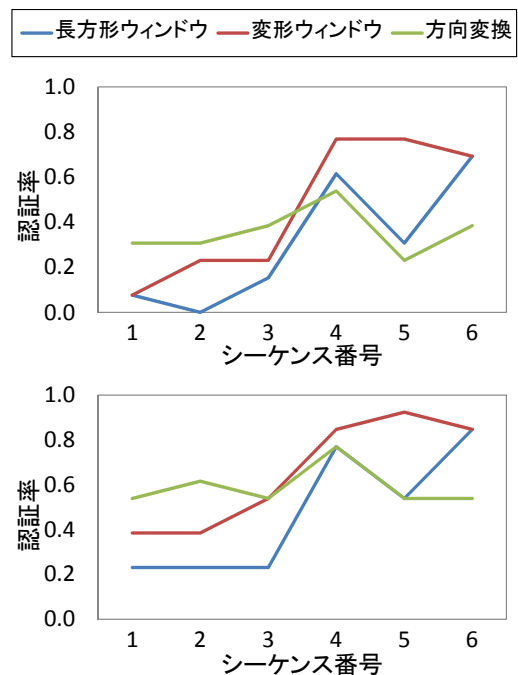


図 9 1 位認証率 (上), 3 位認証率 (下)

Fig. 9 Rank-1 identification rate(upper), Rank-3 identification rate(lower).

8.2 性能評価

カメラ姿勢等を考慮しない通常の長方形のウィンドウで方向変換も行わない場合、カメラ姿勢を考慮したウィンドウで方向変換は行わない場合、カメラ姿勢を考慮したウィンドウで方向変換も行う場合の 3 つの場合について実験を行った。

1 位認証率と 3 位認証率を図 9 に示す。ウィンドウの変形により、長方形のウィンドウより全体的に認証率が向上していることが分かる。一方、方向変換を行うと、観測方向が大きく異なるシーケンス 1~3 は精度が向上している一方、観測方向が類似しているシーケンス 4~6 については、方向変換を用いるとかえって認証率が低下してしまっていることが分かる。方向変換は学習によりある観測角に特徴を近似させるものであるため、ある程度元々の観測角が近ければ、逆に認証率を下げる方向に働いているものと

思われる。

9. おわりに

本論文では、エッジの勾配に着目した特徴量であるSTHOG特徴を用い、その特徴に方向変換モデルを適用させることによって方向変化に頑強な複数カメラ間での人物の照合について述べた。歩行方向を考慮して3次元的に取得した歩行者ウィンドウから特徴を取得したのち、方向の変化による特徴変化を特異値分解により学習してモデルを作成し、新たな歩行者にはそのモデルを適用させることで観測方向を揃えて照合する。照合実験では、大学構内の映像を用い、視点、色合い、背景がそれぞれ異なる場合において、方向変換への対策を行わない場合との比較を示した。

照合実験の結果、カメラ姿勢を考慮したウィンドウの導入により、認証率が向上することが分かった。一方で、方向変換モデルを適用する場合、観測角が大きく異なれば力を発揮するが、観測角の違いが小さい場合は逆に精度を下げる可能性があることが分かった。

今後の課題としては、色特徴のように、比較的方向変化に頑強な特徴とSTHOG特徴を組み合わせて用いることにより、より高い認証率を目指すこと、また、より大きな公開データベースを用いて実験を行うことが挙げられる。

参考文献

- [1] Nakajima, C., Pontil, M., Heisele, B. and Poggio, T.: Fullbody Person Recognition System, *Pattern Recognition*, Vol. 36, No. 9, pp. 1997–2006 (2003).
- [2] 赤塚久哉, 高橋友和, 井手一郎, 村瀬 洋: 不連続な撮影領域間の人物軌跡の対応付けに関する検討, 電子情報通信学会総合大会講演論文集, Vol. 2, p. 198 (2006).
- [3] 本田拓也, 出口大輔, 高橋友和, 井手一郎, 村瀬 洋: 入退室解析のための色特徴を用いた人物対応付けに関する検討, 電子情報通信学会総合大会講演論文集, p. 225 (2009).
- [4] Cheng, E. and Piccardi, M.: Matching of Objects Moving Across Disjoint Cameras, *Proc. of the 18th Int. Conf. on Pattern Recognition*, pp. 1769–1772 (2006).
- [5] Cai, Y. and Pietek, M.: Person Re-identification Based on Global Color Context, *Proc. of the 10th Int. Workshop on Visual Surveillance*, pp. 205–215 (2010).
- [6] Bird, N., Masoud, O., Papanikolopoulos, N. and Isaacs, A.: Detection of Loitering Individuals in Public Transportation Areas, *IEEE Trans. Intelligence Transportation System*, Vol. 6, No. 2, pp. 167–177 (2005).
- [7] Lin, Z. and Davis, L.: Learning Pairwise Dissimilarity Profiles for Appearance Recognition in Visual Surveillance, *Proc. of the 4th Int. Symposium on Visual Computing*, pp. 23–34 (2008).
- [8] Porikli, F.: Inter-camera Color Calibration by Correlation Model Function, *2003 IEEE Int. Conf. on Image Processing*, Vol. 3, pp. II–133–6 (2003).
- [9] Hamdoun, O., Moutarde, F., Stanculescu, B. and Steux, B.: Person Re-identification in Multi-Camera System by Signature based Interest Point Descriptors Collected on Short Video Sequences, *Proc. of the 2nd ACM/IEEE Int. Conf. on Distributed Smart Cameras*, pp. 1–6 (2008).
- [10] Bak, S., E.Corvee, Bremond, F. and Thonnat, M.: Person Re-Identification Using Haar-based and DCD-based Signature, *Prof. of the 7th IEEE Int. Conf. on Advanced Video and Signal Based Surveillance*, pp. 1–8 (2010).
- [11] Viola, P. and Jones, M.: Rapid Object Detection Using a Boosted Cascade of Simple Features, *Proc. of of IEEE computer society conference on Computer Vision and Pattern Recognition 2001* (2001).
- [12] Wang, X., Doretto, G., Sebastian, T., Rittschcer, J. and Tu, P.: Shape and Appearance Context Modeling, *Proc. of the 11th Int. Conf. on Computer Vision*, pp. pp.1–8 (2007).
- [13] Alahi, A., Vanderghyest, P., Bierlaire, M. and Kunt, M.: Cascade of Descriptors to Detect and Track Objects Across Any Network of Cameras, *Computer Vision and Image Understanding*, Vol. 114, No. 6, pp. 624–640 (2010).
- [14] Bak, S., Corvee, E., Beremond, F. and Thonnat, M.: Person Re-identification Using Spatial Covariance Regions of Human Body Parts, *Prof. of the 7th IEEE Int. Conf. on Advanced Video and Signal Based Surveillance*, pp. 435–440 (2010).
- [15] 上村和広, 池亀幸久, 下山 功, 玉木 徹, 山本正信: ネットワーク上の複数カメラを用いた実時間人物照合システム, 信学技報 (PRMU), Vol. 103, No. 659, pp. 67–72 (2004).
- [16] Gheissari, N., Sebastian, T. B., Tu, P. H., Rittscher, J. and Hartley, R.: Person Reidentification Using Spatiotemporal Appearance, *Proc. of of IEEE computer society conference on Computer Vision and Pattern Recognition 2006*, pp. 1528–1535 (2006).
- [17] Berdugo, G., Soceanu, O., Moshe, Y., D.Rudou and Dvir, I.: Object Reidentification in Real World Scenarios Across Multiple Non-overlapping Cameras, *Proc. of the 18th European Signal Processing Conf.*, pp. 1806–1810 (2010).
- [18] Farenzena, M., Bazzani, L., Perina, A., Murino, V. and Cristani, M.: Person Re-Identification by Symmetry-Driven Accumulation of Local Features, *Proc. of of IEEE computer society conferene on Computer Vision and Pattern Recognition 2010*, San Francisco, CA, USA, pp. 1–8 (2010).
- [19] Bazzani, L., Cristani, M., Perina, A., Farenzena, M. and Murino, V.: Multiple-shot Person Re-identification by HPE signature, *20th International Conference on Pattern Recognition* (2010).
- [20] Prosser, B., Zheng, W.-S., Gong, S. and Xiang, T.: Person Re-Identification by Support Vector Ranking, *Proc. of the 21th British Machine Vision Conf.*, pp. 21.1–11 (2010).
- [21] 井尻善久, Lao, S., 村瀬 洋: 非線形距離指標学習によるカメラ間人物照合, 信学技報 (PRMU), pp. 139–146 (2011).
- [22] Zheng, W.-S., Gong, S. and Xiang, T.: Person Re-identification by Probabilistic Relative Distance Comparison, *Proc. of the 22th IEEE Conf. on Computer Vision and Pattern Recognition* (2011).
- [23] Gandhi, T. and Trivedi, M. M.: Person tracking and re-identification: Introducing Panoramic Appearance Map (PAM) for feature representation, *Machine Vision and Applications*, Vol. 18, No. 3, pp. 207–220 (2007).
- [24] Hua, C., Makihara, Y. and Yagi, Y.: Pedestrian Detection by Combining the Spatio and Temporal Features, *Meeting on Image Recognition and Understand-*

- ing (2010).
- [25] Wang, J. and Yagi, Y.: Adaptive Mean-Shift Tracking with Auxiliary Particles, *IEEE Transactions on Systems, Man and Cybernetics -Part B*, Vol. 39(6), pp. 1578–1589 (2009).
 - [26] Wang, J. and Yagi, Y.: Visual tracking and segmentation using appearance and spatial information of patches, *Proceedings of 2008 IEEE International Conference on Robotics and Automation*, Anchorage, Alaska, USA (2010).
 - [27] Wang, J. and Yagi, Y.: Tracking and segmentation using Min-Cut with consecutive shape priors, *Paladyn. Journal of Behavioral Robotics, Versita, co-published with Springer-Verlag GmbH*, Vol. 1, No. 1, pp. 73–86 (2010).
 - [28] Utsumi, A. and Tetsutani, N.: Adaptation of appearance model for human tracking using geometrical pixel value distributions, *Proc. of the 6th Asian Conf. on Computer Vision*, Vol. 2, pp. 794–799 (2004).
 - [29] Sun, J., Zhang, W., Tang, X. and Shum, H.-Y.: Background Cut, *Proc. of the 9-th European Conf. on Computer Vision*, Vol. 2, pp. 628–641 (2006).
 - [30] Dalal, N. and Triggs, B.: Histograms of Oriented Gradients for Human Detection, *IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 886–893 (2005).
 - [31] 川合 諒, 榎原 靖, 八木康史 : STHOG 特徴を用いた複数カメラ間での人物照合, 情報処理学会研究報告, No. 2011-CVIM-177, 茨木, 情報処理学会 (2011).